



# A propos d'une méta-analyse de l'effet de l'agriculture de conservation sur les rendements de diverses cultures en Afrique sub-saharienne

Philippe Letourmy, Marc Corbeels et Krishna Naudin  
(UR Aïda)

10 juillet 2018

## Définition et origine de la méta-analyse

- Une **méta-analyse** est une analyse statistique qui combine les résultats de plusieurs études scientifiques.
  - En agronomie, elle espère élaborer des lois générales sur le fonctionnement des agroécosystèmes, basées sur le traitement statistique d'une base de données, le plus souvent expérimentales, issues des publications retenues.
- La méta-analyse est ancienne (en statistique le premier est [Karl Pearson en 1904](#) sur 5 essais cliniques): la dénomination est de [Gene Glass en 1976](#), ensuite elle s'est développée en santé humaine surtout, mais elle a souvent été cantonnée à de petits regroupements d'études. Dans notre situation d'agronomie globale (ici en Afrique) il s'agit d'une masse plus importante de données, donc un gros volume de travail. Le travail sur la qualité des méta-analyses par [Philibert et al. \(2012\)](#) est très utile.

En agronomie on a beaucoup considéré les regroupements d'essais, souvent en réseaux conçus pour cela. Les premiers travaux en agronomie sont dus à Fisher (1935) et à Yates et Cochran (1938).

Par ailleurs le travail de Philibert, Loyce et Makowski (2012) sur la qualité des méta-analyses est essentiel, me semble-t-il.

## A propos de la méta-analyse utilisée

- Une comparaison entre l'agriculture de conservation et le labour traditionnel a été faite sur la base de résultats de la biblio
- Le présent exposé concerne certains aspects de la procédure d'analyse employée et quelques résultats
- Nous ne nous appesantissons pas sur les multiples difficultés rencontrées et certains aspects méthodologiques ont été déjà présentés le 8/2/2018
- Ce travail est en cours de publication

# 1 Biblio : Agriculture de conservation vs Labour traditionnel

- Recherche de références bibliographiques

- Essais au champ avec au moins une modalité sans labour (voire labour minimum)
- Sous conditions pluviales
- Moyennes, tailles d'échantillons, écarts-type des rendements des cultures accessibles, ou pouvant être calculés à partir des données publiées.

→ 64 articles ont été retenus pour le fichier analysé  
avec 668 observations indépendantes

Cette sélection de références a été conduite par Marc Corbeels et Krishna Naudin, selon une procédure classique en méta-analyse.

## 2 Les traitements de CA vs CT

- Chaque observation du fichier établi pour ces 64 références correspond à la comparaison entre les parcelles CT et les parcelles CA, pour un essai (une année et un lieu) et un traitement (appelé CA\_Type).
- 4 traitements ont été identifiés : NT-M-IR, NT+M-IR, NT-M+IR et NT+M+IR. NT = no tillage, M = mulch, IR = intercropping ou rotation. Nous disposons de 668 observations en tout dans notre fichier

Les références retenues concernaient plusieurs cultures différentes, bien que le maïs représentât plus de la moitié des observations.

### 3 Variable analysée : log ratio

- Soit  $y$  la variable analysée par observation,  $y = \log\left(\frac{CA}{CT}\right)$
- Pour le poids des observations, le choix retenu a été de prendre des poids inversement proportionnels à la variance du log ratio (calculables sur 600 observations sur 668)

CA = rendement en agriculture de conservation

CT = rendement en labour traditionnel

V = variance résiduelle de l'étude divisée par le nombre de répétitions pour la calculer

Un résultat classique de probabilité donne l'approximation :  $Var(f(Z)) = Var(Z) \times (f'(Z))^2$

Ce qui implique :  $Var\left(\log\left(\frac{CA}{CT}\right)\right) = V \times \left(\frac{1}{CT^2} + \frac{1}{CA^2}\right)$ , avec  $V=Var(CT)=Var(CA)$ , constatées identiques

Deux autres choix étaient aussi possibles pour la pondération des données : le poids égal partout et la pondération par les effectifs seulement. Ils n'ont pas été retenus, du fait des différences importantes de variance entre observations, mais aussi des différences importantes dans les nombres d'observations (de  $n=2$  à  $n=176$ , cas d'un essai en milieu paysan). Les distorsions étaient trop importantes pour ne pas essayer d'estimer la précision du log ratio pour chaque observation.

Des auteurs calculent aussi explicitement cette pondération du log ratio : Hedges et al. (1999).

## 4 Sélection des observations

- Le premier problème rencontré a été celui des cultures en trop faibles effectifs, seules les 6 cultures suivantes ont été retenues : « cotton », « cowpea », « maize », « rice », « sorghum » et « soybean »
- Le deuxième problème concernait les covariables explicatives : c'était la difficulté à attribuer une valeur identique de « applied\_N » ou de « applied\_P » à CT et à CA d'une même observation
- Tous ces éléments nous ont conduits à retenir 549 observations en tout pour cette méta-analyse

Cowpea = niébé

## 5 Modèle mixte simple de regroupement

- Le log ratio a été étudié en fonction des traitements utilisés et des cultures, tout en tenant compte des effets des essais et de l'interaction traitement-essai
- $Y_{ijkl} = \mu + \alpha_i + \beta_{ij} + C_{ik} + D_{ijk} + \varepsilon_{ijkl}$
- Y étant le log ratio, pour la culture i, le traitement j, l'essai k et l'observation l;  $\mu$  la moyenne générale,  $\alpha_i$  l'effet fixe de la culture,  $\beta_{ij}$  l'effet fixe du traitement (CA\_Type) j et la culture i,  $C_{ik}$  est l'effet aléatoire de l'essai k et la culture i,  $D_{ijk}$  l'effet aléatoire d'interaction entre le traitement et l'essai k pour la culture i, et  $\varepsilon_{ijkl}$  l'erreur résiduelle
- Nous avons appelé essai un ensemble d'observations avec des traitements CA différents, mais un même témoin CT

La variable essai a été appelée Id1 dans notre fichier (elle a dû être entrée manuellement dans le fichier).

Les résultats obtenus pour ce modèle simple ont déjà été présentés le 8/2/2018.



## 6 Modèle mixte avec covariables

- Les résultats du modèle simple sur les effets culture et traitement ne tenaient pas compte des conditions extrêmement variées des essais compilés : le climat, le type de sol, la pluviométrie..., mais aussi les itinéraires techniques
- L'idée est alors d'estimer ces effets fixes en essayant de se placer « en conditions identiques », c'est-à-dire en introduisant une liste de covariables disponibles sur chacune des publications
- L'espoir est aussi de diminuer la variance résiduelle du modèle, mais aussi l'interaction essai-traitement, voire l'effet essai (effets aléatoires)

## Modèle mixte avec covariables (suite)

- Estimer ces effets culture et d'interaction traitement-culture, tout en tenant compte de variables d'intérêt
- $Y = \mu + \alpha + \beta + \sum_m (X_m) + C + D + \varepsilon$
- Y étant le log ratio,  $\mu$  la moyenne générale,  $\alpha$  l'effet fixe de la culture,  $\beta$  l'effet fixe du traitement (CA\_Type) par culture, C est l'effet aléatoire de l'essai (pour chaque culture), D l'effet aléatoire d'interaction entre le traitement et l'essai, par culture, et  $\varepsilon$  l'erreur résiduelle
- Les variables  $X_m$  sont les covariables explicatives introduites; elles ont été découpées en classes (devenues des variables qualitatives ici)
- La liste des covariables a été établie « à dire d'expert » et en tenant compte des informations disponibles dans les publications

## Modèle avec covariables (suite)

- Les covariables initialement identifiées étaient: climat\_koppen, soil\_nrccs, study\_type, rainfall\_type, herbicide\_on\_ca, average\_rainfall, applied\_N, applied\_P, CT\_Type (plow, hoe, etc.), trav\_sol (no-tillage, minimum-tillage, basins), Altitude, Initial\_soil\_C, Average\_Temp, Legume\_rotation, Legume\_intercropping, Initial\_mulch, Mulch\_type
- Cela ne marchait pas ! Deux problèmes : les données manquantes et les colinéarités
- Le C initial du sol a été mis de côté (données manquantes)
- La sélection a été refaite en principalement sur les effets de colinéarités (et très secondairement sur des  $F < 1$ : effets « négligeables »)

Problème de la procédure de sélection des covariables, une fois écartée la question des colinéarités: sélection d'effets fixes ou aléatoires dans un modèle mixte

Référence: *Zuur et al. (2009). Mixed Effect Models and Extensions in Ecology with R.* Ils recommandent une stratégie descendante comme suit:

1. adapter un modèle complet (aller même "au-delà de l'optimum", c'est-à-dire plus complexe que ce qu'on attend)
2. trier la structure des effets aléatoires (utiliser les vraisemblances REML, le cas échéant avec AIC ou BIC)
3. trier la structure des effets fixes (soit utiliser REML la statistique F, soit comparer les modèles ML imbriqués – en gardant les effets aléatoires constants)
4. une fois arrivé au modèle final, utiliser l'estimation REML

## Modèle utilisé à ce jour

- Finalement les covariables choisies étaient : climat\_koppen, soil\_nracs (structurée ou non en texture et reste de l'effet sol), study\_type, trav\_sol, herbicide\_on\_ca, a\_rain et ap\_N
- Les variables average\_rainfall et applied\_N ont été mises en classes (notées a\_rain et ap\_n)
- Alors se pose aussi 2 problèmes :
  - 1 le calcul des moyennes ajustées pondérées (sont-elles estimables ?)
  - 2 le calcul de la transformée inverse, car nous avons les moyennes en  $\log(\text{CA}/\text{CT})$  au lieu du rapport CA/CT ou d'une différence entre CA et CT

La forte corrélation entre certaines covariables peut rendre les différentes moyennes ajustées non estimables: c'est une cause possible de simplification du modèle. Le calcul des moyennes ajustées doit être fait en tenant compte des poids choisis des observations, et il faut reprendre les sorties sous Excel pour faire les calculs de transformée inverse et en déduire les intervalles de confiance (c'est un gros travail).

# Analyse de variance du modèle avec covariables

Valeurs estimées des paramètres de covariance	
Param de cov	Valeur estimée
Id1	0.007280
Crop*CA_type*Id1	0.005328
Residual	0.001254

Type 2 Tests des effets fixes				
Effet	DDL Num.	DDL Res.	Valeur F	Pr > F
Crop	5	91	1.66	0.1512
Crop*CA_type	11	91	2.38	0.0124
Climat_Kop*Soil_NRCS	18	89	3.18	0.0002
ap_n	3	89	0.97	0.4111
Study_type	1	89	21.52	<.0001
a_rain	3	89	4.26	0.0074
Herbicide_on_CA	2	89	10.58	<.0001
trav_sol	2	89	6.99	0.0015

Par rapport aux résultats bruts, l’effet du traitement est moins significatif, mais il y a un effet type d’étude et un effet herbicide à noter.

# Résultats du calcul des moyennes maïs par traitement (modèle avec covariables)

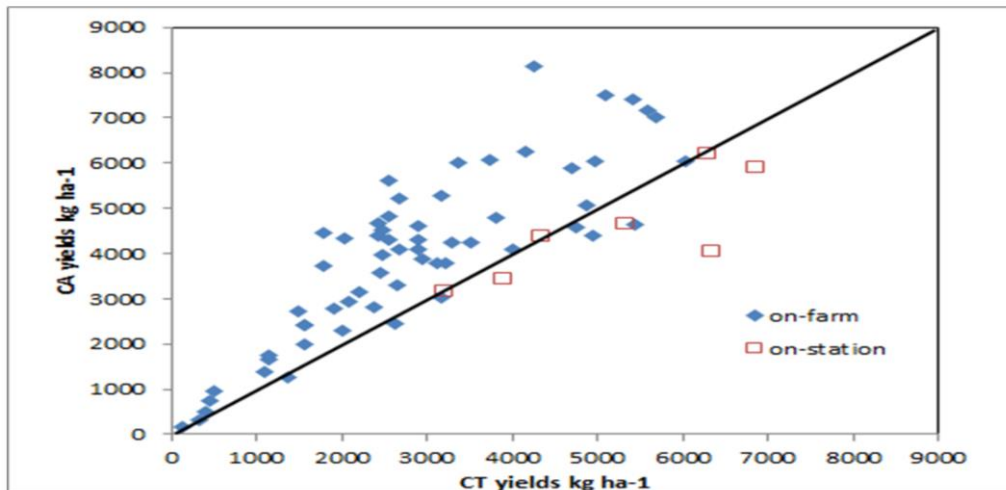
Effet	Crop	CA_type	Moyenne	min%	moy%	max%	
Crop*CA_type	maize	NT+M+IR	0,01296	96,19	101,30	106,69	
Crop*CA_type	maize	NT+M-IR	0,03532	100,71	103,60	106,56	CA>CT
Crop*CA_type	maize	NT-M-IR	0,02180	97,66	102,20	106,96	
	Study_type						
Study_type	on-farm		0,09647	106,34	110,13	114,05	CA>CT
Study_type	on-station		-0,02780	94,36	97,26	100,24	
	Herbicide_on_CA						
Herbicide_on_CA	N		-0,00444	97,12	99,56	102,05	
Herbicide_on_CA	Y		0,06162	103,78	106,36	109,00	CA>CT

Les moyennes présentées sont celles du log ratio ! Et la transformation inverse donne le rapport CA/CT en pourcentage.  
 Par rapport aux résultats bruts, l’effet du traitement, sur le maïs en particulier, est beaucoup moins fort.

## Modèle avec covariables

- Des variables avec beaucoup de données manquantes, comme le C initial du sol, ont été réintroduites dans le modèle, mais sans succès
- On aboutit à des résultats sensiblement différents du modèle simple (en valeurs moyennes et en comparaisons): comment peut-on interpréter ces divergences sur le plan agronomique ?

## Type d'étude - maïs (traitement NT+M+IR)

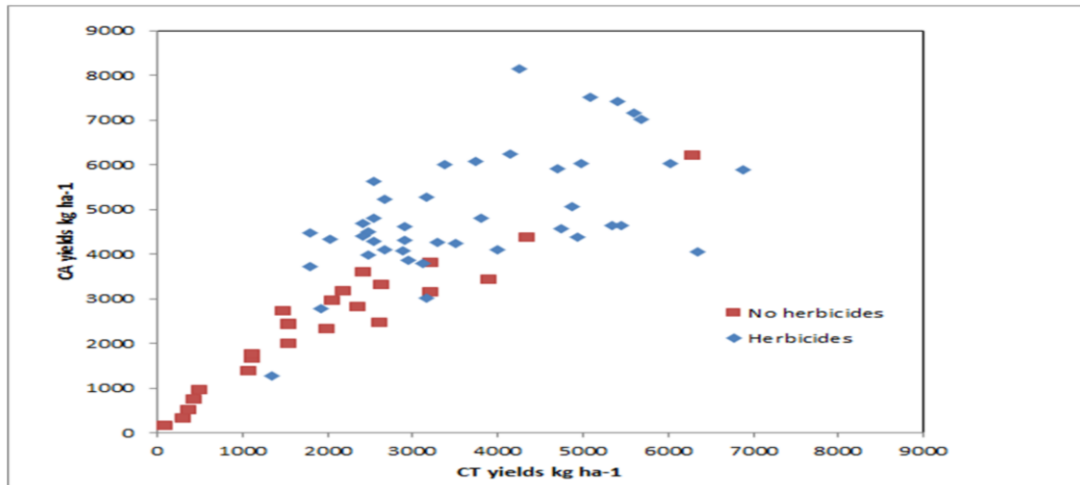


Il n'y a que le traitement CA (NT+M+IR) et il ne s'agit que du maïs.

Plusieurs interprétations sont possibles: ou bien il y a un biais de sélection des parcelles d'essais en milieu paysan (situations défavorables d'essais favorisant l'effet de l'agriculture de conservation, ou bien on suppose une moindre maîtrise des techniques d'agriculture de conservation par les responsables de culture en station qu'en milieu paysan.



## Effet herbicide - maïs (traitement NT+M+IR)



Il n'y a que le traitement CA (NT+M+IR) et il ne s'agit que du maïs.  
Sur les mêmes données, on constate un effet relativement plus important de CA contre CT lorsqu'on utilise de l'herbicide que lorsqu'on en n'utilise pas.

## Discussion et conclusions

- La démarche a été bien établie (la **pondération** utilisée est fondée sur le calcul, la procédure de **sélection de variables** de Zuur et al. (2009) a semblé utilisable
- les résultats montrent des effets du type d'étude et de l'herbicide qui se substituent en partie à celui des traitements étudiés. Il reste à formuler une **interprétation agronomique** de ces effets
- Il reste quelques points à éclaircir : y a-t-il un moyen de mieux exploiter les variables ayant des **données manquantes** et quelle mesure de **sensibilité** des covariables entrées ?

Merci de votre attention !

#### Références

- 1) Aurore Philibert, Chantal Loyce, David Makowski. Assessment of the quality of meta-analysis in agronomy. *Agriculture, Ecosystems and Environment*, Elsevier Masson, 2012, 148, pp.72 - 82. [〈10.1016/j.agee.2011.12.003〉](#) . [〈hal-01004278〉](#)
- 2) Zuur, A., Ieno, E.N., Walker, N., Saveliev, A.A., Smith, G.M. (2009). *Mixed Effects Models and Extensions in Ecology with R*. Springer-Verlag New York. XXII, 574 p.